

# *Counting Reads and Differential Expression*

Short Read Workshop  
Day 7

# Recap of Day 7 Videos

- Counting reads (for RNA-seq)
  - Steps to counting reads and normalizing counts
- Differential expression overview
  - Introduction to differential expression analysis
- Differential expression analysis with DESeq2
  - DESeq2 Negative Binomial theoretical model
- Multifactor designs with DESeq2
  - Expanded designs in DESeq2

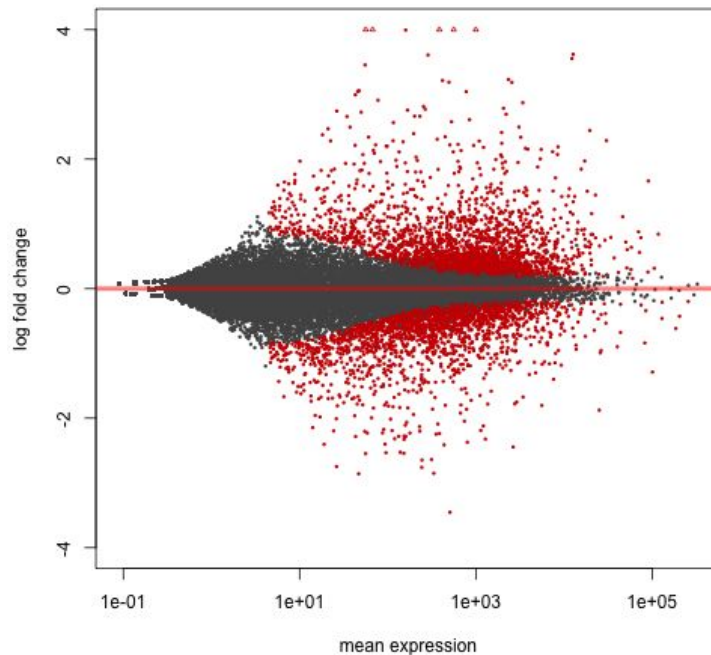
# Goal of the Day

Find genes that are different between samples

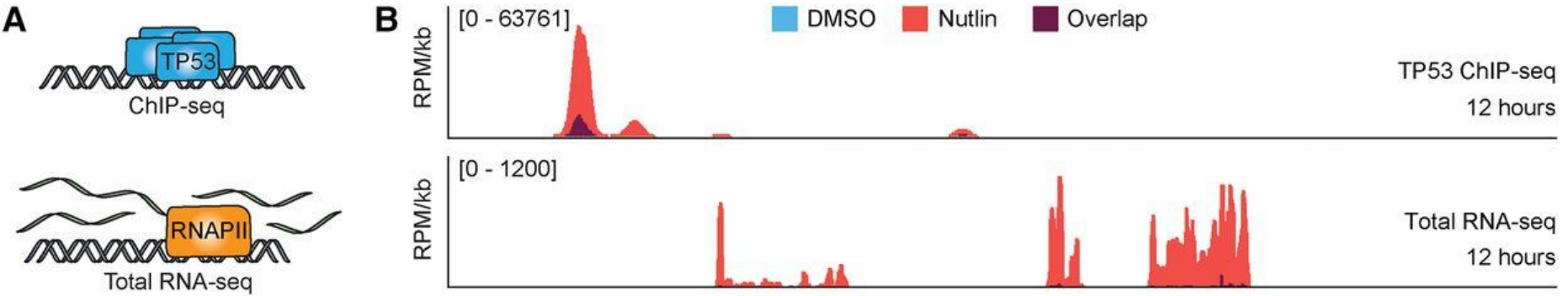
Map reads to reference genome

Count reads

Perform differential gene  
expression analysis



# Project B: Identification of the p53 transcriptional program using RNA-seq and ChIP-seq



Andrysik et al., 2017, doi: [10.1101/gr.220533.117](https://doi.org/10.1101/gr.220533.117)

In HCT116, SJSA, and MCF7 cell types.

Question: Which genes are driven by p53 activation?

# featureCounts counts reads over features in R

There are several options in featureCounts

```
fc <- featureCounts(files=bam_file_list,
  annot.ext=gtf,
  isGTFAnnotationFile=TRUE,
  GTF.featureType="exon",
  GTF.attrType="gene_id",
  useMetaFeatures=TRUE,
  allowMultiOverlap=TRUE,
  largestOverlap=TRUE,
  countMultiMappingReads=TRUE,
  isPairedEnd=TRUE,
  strandSpecific=1,
  nthreads=N)
```

|  | union   | intersection_strict | intersection_nonempty |
|--|---|---------------------|-----------------------|
|  | gene_A  | gene_A              | gene_A                |
|  | gene_A  | no_feature          | gene_A                |
|  | gene_A  | no_feature          | gene_A                |
|  | gene_A  | gene_A              | gene_A                |
|  | gene_A  | gene_A              | gene_A                |
|  | ambiguous<br>(both genes with --nonunique all)            | gene_A              | gene_A                |
|  | ambiguous<br>(both genes with --nonunique all)            |                     |                       |
|  | alignment_not_unique<br>(both genes with --nonunique all) |                     |                       |

aligned read:  
start: 113217600 end: 113217650



GTF

|      |         |             |           |           |   |   |   |
|------|---------|-------------|-----------|-----------|---|---|---|
| chr1 | unknown | exon        | 113217048 | 113217252 | - | + | gene_id "MOV10";p_id "P5535";transcript_id "NM_001130079" |
| chr1 | unknown | exon        | 113217048 | 113217351 | - | + | gene_id "MOV10";p_id "P5535";transcript_id "NM_020963"    |
| chr1 | unknown | exon        | 113217470 | 113217671 | - | + | gene_id "MOV10";p_id "P5535";transcript_id "NM_001130079" |
| chr1 | unknown | CDS         | 113217535 | 113217671 | - | + | gene_id "MOV10";p_id "P5535";transcript_id "NM_001130079" |
| chr1 | unknown | start_codon | 113217535 | 113217537 | - | + | gene_id "MOV10";p_id "P5535";transcript_id "NM_001130079" |

↑  
feature type

↑  
feature

# Counting reads with featureCounts

- Follow [featureCounts](#) worksheet:
  - Open R and install Rsubread on AWS
  - Get `d7_featureCounts.R` and `d7_featureCounts.sbatch` scripts
  - Edit both scripts and execute the sbatch script

## Challenge Question

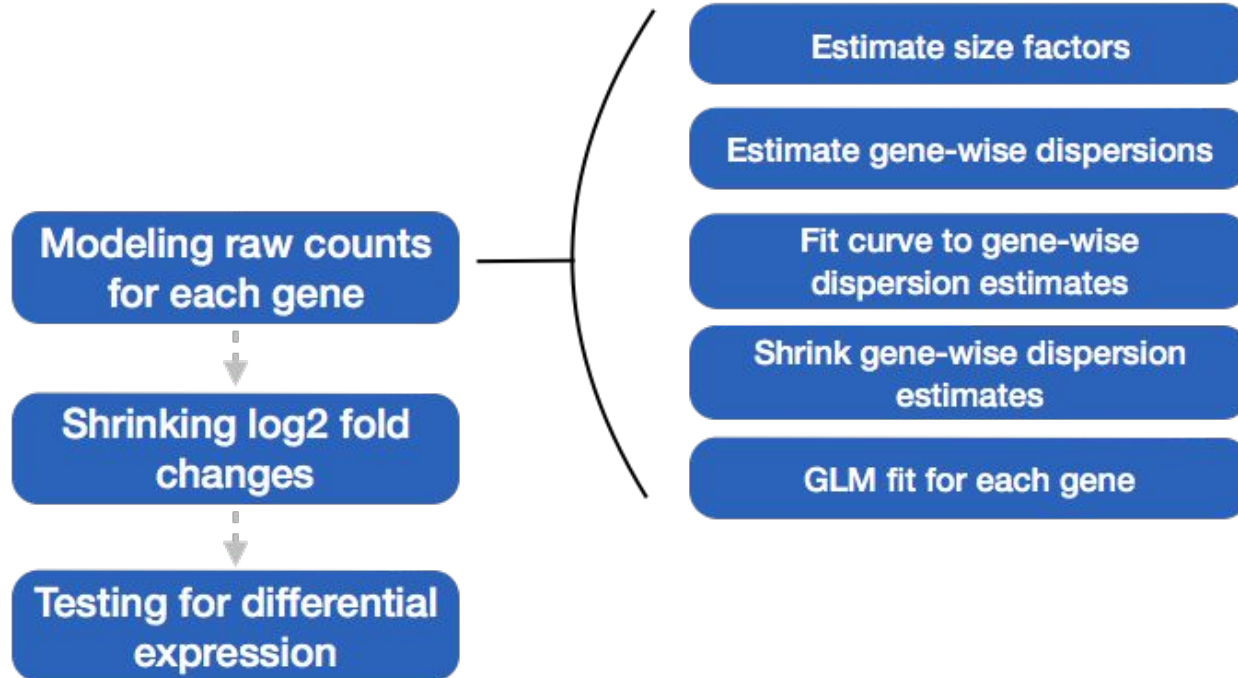
- What feature would you use to count reads for RNA-seq?
- A. Gene
  - B. Exon
  - C. Transcripts

## Challenge Question

- What feature would you use to count reads for RNA-seq?
- A. Gene
- B. Exon ✓
- C. Transcripts

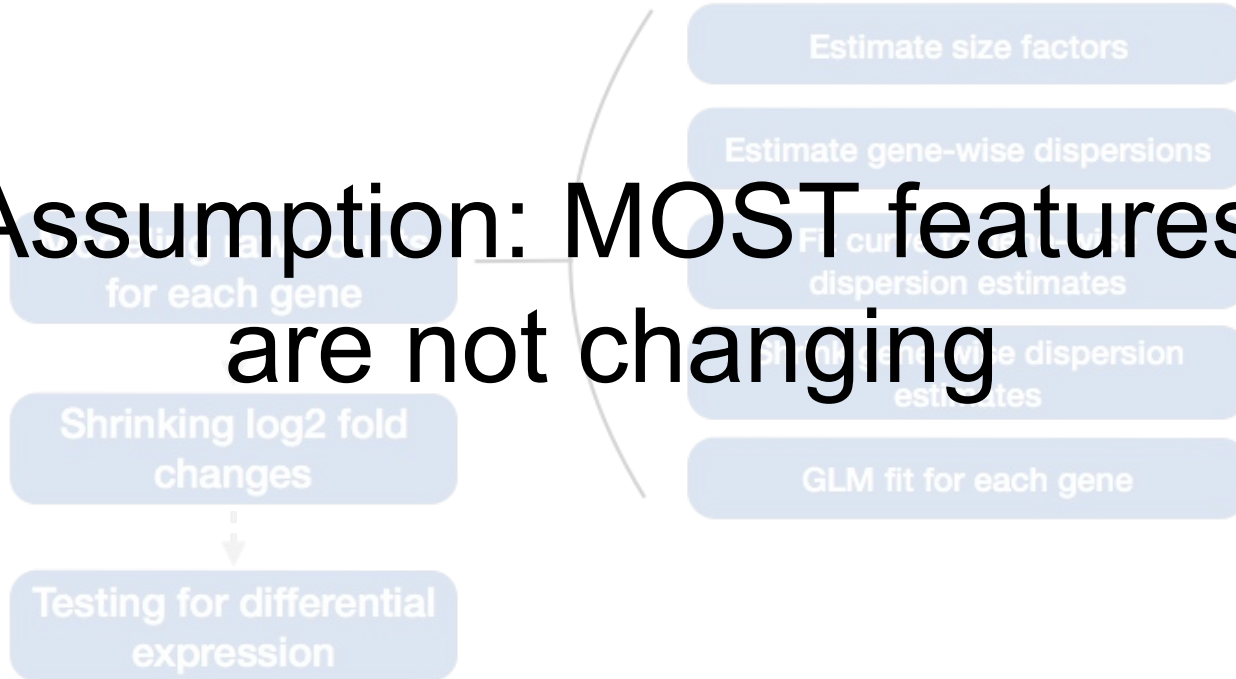


# DESeq2 Recap



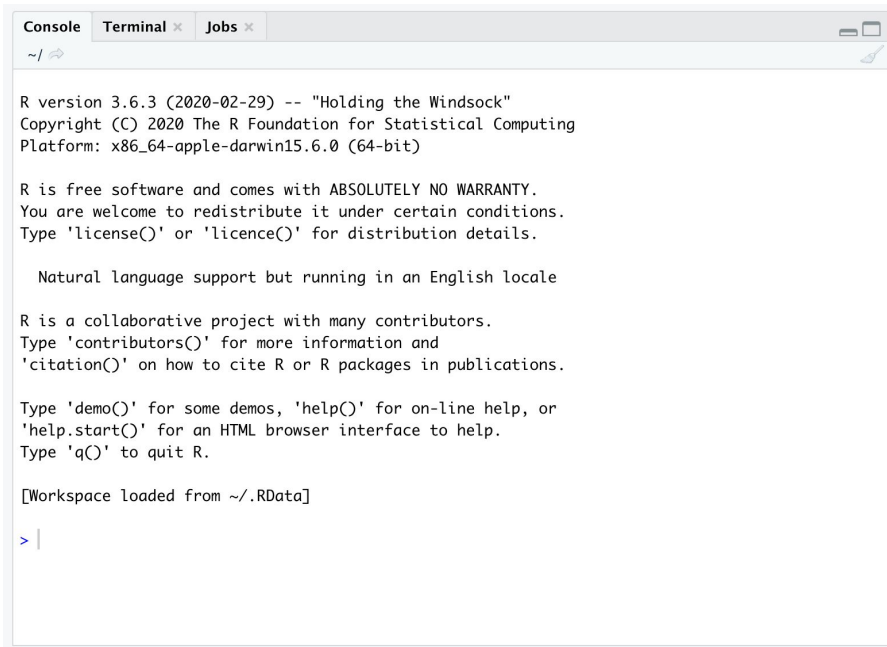
# DESeq2 Recap

Assumption: MOST features  
are not changing



# Run DESeq2...

- Follow [DESeq2](#) worksheet
  - This will be run in an R console in RStudio



```
Console Terminal x Jobs x
~/
R version 3.6.3 (2020-02-29) -- "Holding the Windsock"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/.RData]

> |
```

# Challenge Question

- How would you run DESeq2 on the supercomputer?

# Challenge Question

- How would you run DESeq2 on the supercomputer?
  - Install DESeq2 in your R packages directory
  - Make a conditions table that matches your count table
  - Run the R script through an sbatch script

# Homework

- Run DESeq2 to explore differential expression with a different cell line